1. Data selection
    - How many waves
    - How many variables and sample size
    - Data quality(missing value, data accuracy)
    - Train test split

2. Data cleaning & preprocessing
    - Duplicated samples
    - Missing value imputation
    - Data transformation
    - Normalization

3. Imbalanced problem
    - Up-sampling
    - Down-sampling
    - SMOTE(Synthetic Minority Over-sampling Technique)

4. Model comparison
    - Classification ML model selection
    - Model training(Cross-validation)
    - Hyper-parameter tuning

5. Model evaluation
    - Evaluation metrics (ACC,AUC)
    - unseen data prediction

6. Feature importance
    - List feature importance from tree-based model to create efficient data collection channels

7. Some Limitations might be concerned
    - Insufficient data
    - Inconsistency depression ratio among different age range(depression ratio might be higher in a certain age range than others)
    - Improper risk factors (such as financial crisis, might only be relevant to adult populations but not for all the people who were 18 years and older)